DYNAMIC PIXEL BINNING ALLOWS SPATIAL AND ANGULAR RESOLUTION TRADEOFFS TO IMPROVE IMAGE QUALITY IN X-RAY C-ARM CT

Alexander Steg, Marc Reichenbach, Christopher Söll, Lan Shi, Andreas Maier, Christian Riess

Department of Computer Science, Department of Electrical Engineering
Friedrich-Alexander University Erlangen-Nuremberg,
Germany

ABSTRACT

When designing an acquisition protocol for a C-arm CT, a tradeoff has to be made between scan time and the amount of acquired data. Increased scan time allows data acquisition with higher spatial and angular resolution. Fixed scan time implies a reduction of either spatial and angular resolution, which may either lead to oversmoothing or to artifacts from angular undersampling.

In this work, we investigate smart sensors for pushing the boundary for this bottleneck considerably further. Detector pixels are enhanced by processing units that are able to perform a limited amount of local operations. This allows an online, dynamical adaptation of the pixel binning, depending on whether a local neighborhood is smooth or not. We demonstrate on phantom data that this approach allows to save about 75% of the data, obtaining a reconstruction quality comparable to 1×1 binning at a data rate comparable to 2×2 pixel binning.

***Index Terms*—** Pixel Binning, X-ray detector, C-arm reconstruction, Dynamic Pixel Binning, X-ray C-arm CT

1. INTRODUCTION

An ongoing challenge in X-ray CT reconstruction is to compute high-quality, sharp anatomic volumes. Patient motion during the scan can severely distort the reconstruction result [1]. Thus, manufacturers aim to keep the scan time at a minimum. In practice, particularly for C-arm CT systems, the scanning protocol is expedited by limiting the transfer of the data, such that a volume scan can be finished within a few seconds. This data limitation can either be realized a) by reducing the pixel resolution at the detector via binning, or b) by acquiring a lower number of projections, which comes at the expense of a decreased tomographic angular sampling, and hence stronger artifacts in the reconstruction. To illustrate this with numbers, a typical C-arm CT scan protocol is to acquire 493 projections with $1024 \cdot 768$ pixels resolution. An alternative is to use very coarse angular sampling, e. g. 123 projections with a much higher image resolution of $2048 \cdot 1536$ pixels. Thus, by a back-of-the-envelope calculation, one may reasonably speculate that the maximum data transfer of such a system is about 800MB per scan (by using, e. g., the latter configuration, and computing $123 \cdot 2048 \cdot 1536 \cdot 2\,\mathrm{byte} = 773,849,088\,\mathrm{byte}$).

We build on recent advances in detector technology to explore a new approach to tradeoff angular and spatial resolution. The key concept to reduce the data amount is to process part of the information right after signal acquisition, as described in [2]. This leads to the concept of smart cameras or smart sensors, meaning processing hardware placed close to the image sensor. By using a smart camera, a hardware device is connected directly to the data stream, to process the incoming information and reduce the amount of data for transfer. A good overview about the major development in the smart camera topic is presented in [3]. Moreover, a prototype of such a camera in the medical environment has been displayed in [4].

In this work, we demonstrate that performing a limited set of *pixel-local* operations on the detector makes it possible to employ a dynamic binning scheme. Strongly varying areas are stored with 1×1 pixel binning, while mostly homogeneous areas are stored with a binning of 2×2 pixels or larger. We demonstrate in a proof-of-principle investigation the feasibility of obtaining reconstructions of comparable quality with a much reduced amount of data. These findings can be further developed to either relieve the data bottleneck for allowing faster scanning protocols, or to improve reconstruction quality with time- and data budgets similar to current protocols.

2. RELATED WORK

There exist already strategies to limit scan time and data volume on C-arm CT systems while maintaining high-quality reconstructions. Long scan times may lead to (involuntary) patient motion or organ motion during acquisition. For such situations, one may use a data-driven compensation of motion artifacts after the acquisition [5]. One approach to faster acquisitions is angular projection undersampling, together with iterative reconstruction [6, 7]. Here, radiation dose is the main motivation for reducing the number of projections. The main drawback of such approaches is that the reconstruction is computationally an order of magnitude more complex. Furthermore these algorithms rely on relatively strict smoothness assumptions.

X-ray detectors in current C-arm CT products are operated with a fixed binning size, for instance 1×1, 2×2, or 4×4 pixel binning, which is set prior to acquisition. However, there are several groups working on more intelligent detectors. Here, logic circuits are added on top of one or more neighbored pixels, such that the detector, acting as a smart camera, can perform a limited set of computational operations. Smart camera applications have been presented in various prior works. Astrom *et al.* developed an image sensor which is able to process local operations, such as gradient calculation and median filtering [8]. For a more coarse-grained processing, Albani *et al.* presented presented the integration of a RISC processor, as well as parallel co-processor, on the image sensor die to allow arbitrary image processing operations [9]. Moreover, high speed smart sensors have been developed as shown by Dubois *et al.* [10]. They built a 10000 fps image sensor with preprocessing capabilities such as edge detection, which can be used for pixel binning among other operations. One general limitation of data processing directly on the detector is that almost all operations have to be spatially local.

3. METHODS

We propose a sequence of local operations to dynamically adapt the detector binning. We are seeking a tradeoff between artifacts from spatial and angular undersampling. The proposed method preserves high spatial resolution in non-smooth areas exhibiting high frequency components by applying a 1×1 pixel binning. We reduce the spatial resolution in smooth areas (i. e., where a high resolution is less critical) by applying larger binning. As a result, the reconstructed volume provides better image quality than uniform 2×2 binning projections with a comparable amount of data.

3.1. Calculation of Dynamic Bins

When a projection image is acquired, each pixel creates a local copy of the captured intensity. We subdivide the pixels into a grid of $N \times N$ pixels cells, and use the local copy to determine whether the pixels in a grid cell neighborhood are smooth or non-smooth. Upon detector readout, we report for non-smooth areas the exact original pixel value (i. e., perform local 1×1 binning). If the area is smooth, the pixel becomes part of a $N \times N$ bin, and only an average value of the neighborhood is reported (corresponding to a binning of 2×2, 3×3, or more pixels).

To determine whether the neighborhood is smooth, we first apply bilateral filter on the copied pixels to remove small amounts of noise [11]. Then, we compute on this filtered image the edges using a Sobel filter to enhance structural variations. The smoothness of the area is determined by computing the standard deviation of the edge map. If the standard deviation is below a threshold t, the neighborhood is assumed to be smooth. An important property of all these operations is that none of them requires global information, and can be readily built in hardware [11].

The data reduction is achieved by encoding each of the averaged pixels only once. For example, if a pixel is part of a 3×3 pixels bin, its net storage size is less than 15% of a 1×1 binned pixel.

3.2. Smoothness Information in Consecutive Projections

Given that the filtering steps on the detector require a short, but measurable amount of time, we implement the smoothness test and the data compression in parallel. During tomographic acquisition, we use the smoothness test from the previous projection for encoding the binning of the current projection: if there is a non-smooth pixel in the previous projection, its maximum possible shift in the current projection is bounded by the CT geometry, the angular sampling, and the width of the patient. In a typical setup, this maximum shift is only a few pixels, which are added as an additional safety margin for which the 1×1 binning is also applied.

3.3. Decompression of Dynamically Binned Images

At the workstation, each pixel intensity is restored by decompressing the image buffer. 1×1-binned pixels are exactly restored. A pixel from a $N \times M$-binning region is interpolated from itself and its immediate neighboring pixels using a Gaussian weight mask, to reduce sharp edges and to reduce artifacts in the reconstruction.

4. EXPERIMENTS AND DISCUSSION

A prototype implementation is done in the CONRAD reconstruction framework [12]. We simulate a cone-beam CT scan with a source-to-detector distance of $1200\,\mathrm{mm}$. We use two datasets based on the geometry of the FORBILD head phantom [13]. In contrast to the original FORBILD phantom, our version is rendered with actual material definitions from the NIST database [14]. The first dataset is noise-free, the second dataset is simulated with 50000 photons using a monochromatic energy of $80\,\mathrm{keV}$ for moderate noise, as in [15]. The simulated detector size is 640×480 pixels with a pixel spacing of $1.2\,\mathrm{mm}$ in both x and y direction. The phantom is located exactly in the center between X-ray source and detector. The trajectory rotates around the z-axis with 248 projections, corresponding to an angular increment of $0.869°$. The reconstructed volume consists of $512 \times 512 \times 512$ voxels with a spacing of $0.5\,\mathrm{mm}$. For reconstruction, the FDK cone beam algorithm is used [16], applying a Ram-Lak ramp filter with cosine and Parker weights.

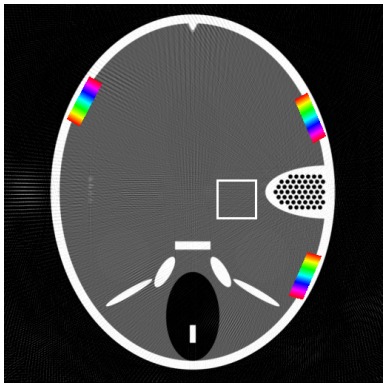


Fig. 1. Line profiles for MTF and ROI for σ_{sd} calculation.

For our experiments we only use quadratic binning masks of 2×2, 3×3 and 5×5 pixels (although one may also think, e. g., of $1 \times N$ or $N \times M$ masks). The bilateral filter operates on 5×5 pixels with a photometric distance of 0.001 and a geometric distance of 2.0 , the Sobel kernel operates on 3×3 pixels. Multiple binning kernel sizes are compared to two reference images, namely the reconstructed 1×1 binning (full resolution) volume and a 2×2 binned volume, as it is the current standard for scans with high angular resolution.

We evaluate several evaluation metrics. We use the structural similarity index (SSIM) [17] to quantify the image quality. Additionally, we compute the standard deviation σ_{sd} in a homogeneous box of $25\,\mathrm{mm}$ side length. The position of the box is shown in Fig. 1. The reduction in the amount of data r is computed as

$$r = 1 - \frac{d}{f} \ , \tag{1}$$

where f is the total amount of data for uniform 1×1 binning, and d is the required data for dynamic binning. For example, uniform 1×1 pixel binning yields 0% saved data, and uniform 2×2 pixels binning yields 75% saved data. For dynamically binned data, each $N \times N$-binned region has one additional bit attached, indicating whether this region is binned. Thus, the theoretical maximum data reduction, e. g., for dynamic 2×2 binning, is 73.43% (we did not investigate bitmask compression). The resolution of the reconstructed central slice is calculated with the modulation transfer function (MTF) taken from a mean line profile of 270 lines across the inner edge of the skull bone as shown in Fig. 1. The achieved resolution is the spatial frequency which corresponds to 10% of the normalized MTF.

Table 1 and Tab. 2 summarize our quantitative results. The columns indicate the size of the binning kernel, the dynamic binning threshold t, the data reduction rate r, SSIM compared to uniform

Bins	t	r	SSIM(1)	SSIM(2)	MTF	σ_{sd}
Uniform						
1×1		0	1	0.839	1.114	0.238
2×2		75	0.839	1	0.702	0.222
Dynamic						
3×3	0.01	75	0.950	0.831	1.118	0.236
2×2	0.03	65	0.982	0.837	1.092	0.236

Table 1. Settings and quantitative results on noise-free data

Bins	t	r	SSIM(1)	SSIM(2)	MTF	σ_{sd}
1×1		0	1	0.773	1.112	0.242
2×2		75	0.773	1	0.712	0.223
3×3	0.07	75	0.893	0.783	1.108	0.239
5×5	0.08	80	0.934	0.781	1.073	0.242

Table 2. Settings and quantitative results on noisy data

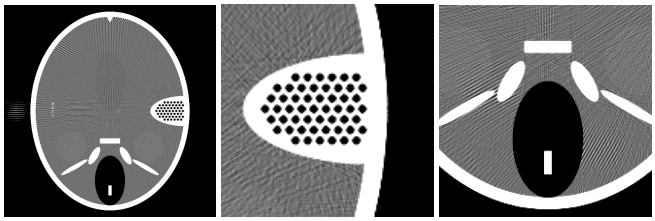
1×1 and 2×2 pixel binning (denoted SSIM(1) and SSIM(2)), the spatial frequency at 10% of the MTF, and the standard deviation σ_{sd}.

Results on the noise-free phantom are shown in Fig. 2. Figure 2 (a) and Fig. 2 (b) show uniform 1×1 and 2×2 binning, respectively. As expected, the reconstruction from uniform 2×2 binning is blurred and shows considerably less details. This is best seen in the inner ear and in the rectangular bone box of the frontal sinus. The visual impression is confirmed by quantitative results in Tab. 1. Here, 2×2 binning exhibits a lower resolution and a lower structural similarity index compared than 1×1 binning.

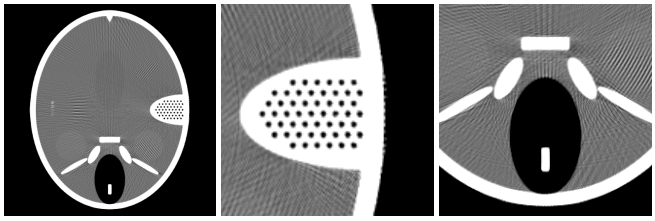
Figure 2 (c) and Fig. 2 (d) illustrate the impact of dynamic binning. Compared to uniform 2×2 binning, much more details are preserved, leading to an overall sharper image. Figure 2 (c) for dynamic 2×2 binning shows that we can almost achieve the quality of the uniform 1×1 binning image. Some artifacts were introduced, which we assume to originate from the undersampling at binned-to-non-binned discontinuities of the decompressed projections. Table 1 quantitatively confirms these observations. Particularly, dynamic 3×3 pixels bins with $t = 0.01$ saves 75% of data, which is the same amount like uniform 2×2 pixel binning. Yet, this image has a much higher SSIM(1) compared to uniform 2×2 binning. Overall, the standard deviation and the 10% MTF are in the same order of magnitude for uniform 1×1 binning and dynamic binning, with a SSIM greater than 95%.

Using the same reconstruction pipeline, Tab. 2 and Fig. 3 show results on the noisy dataset. Analogously to the noise-free dataset, the uniformly 2×2 binned image shows less detail in the inner ear and blurs around sharp edges (cf. Fig. 3 (b)). While keeping the same data saving factor $r = 75\%$ like uniform 2×2 binning, dynamic 3×3 binning increases the SSIM(1) by almost 16% and the resolution by 64%. Due to the image noise, t was set to 0.07. Applying a 5×5 binning, we saved 80% of the data using a threshold $t = 0.08$. However, this comes at the expense of a slightly reduced resolution and a structural similarity of 0.93, which shows in some streaks in the frontal sinus in Fig. 3 (d).

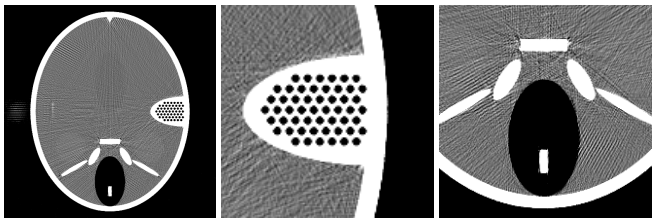
As a general note, the percentage of reduced data strongly depends on the homogeneity of the object. We expect that the data reduction on real data turns out to be considerably lower than on phantom data, which we will investigate in future work. However, we see this work as a proof-of-concept implementation, demonstrating that dynamic pixel binning has the potential to greatly reduce the amount of projection data while providing reconstructed volumes with an excellent level of detail.



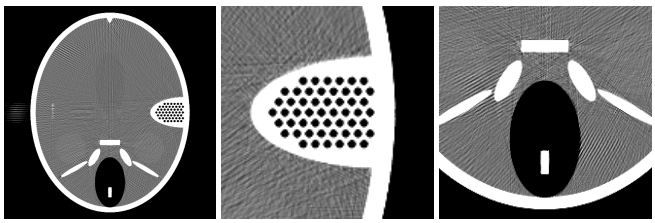
(a) uniform 1×1 binning, $r = 0\%$, SSIM(1)$= 1$



(b) uniform 2×2 binning, $r = 75\%$, SSIM(1) $= 0.839$



(c) dynamic 3×3 binning, $r = 75\%$, SSIM(1) $= 0.950$

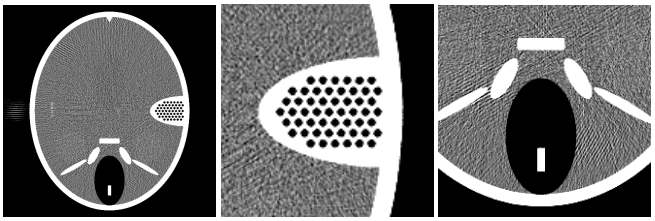


(d) dynamic 2×2 binning, $r = 65\%$, SSIM(1) $= 0.982$

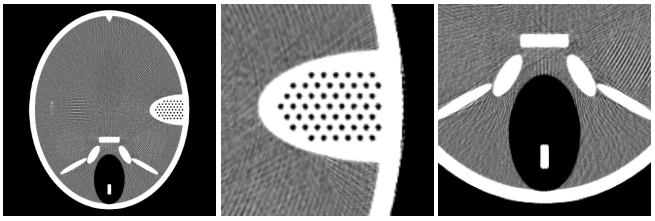
Fig. 2. Qualitative results on noise-free phantom data. Simulation parameters and quantitative metrics are listed in Tab. 1. Window: C=60HU, W=500HU. See text for details.

5. CONCLUSION

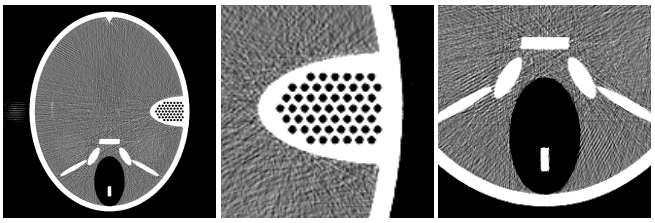
Data transfer times from the detector to the workstation limit the practically achievable angular and spatial resolution of C-arm CT projections. Adaptive pixel binning on smart detector pixels may be a viable way to push the data limit considerably higher. We have shown that an adaptive binning strategy increases the quality of the reconstructed volumes, while saving similar amounts of data as applying uniform 2×2 binning methods. This is achieved by performing spatially local processing operations at the detector itself, leading to dynamic pixel binning for each projection. Results show that the proposed method is able to create high-quality reconstructions on noisy data. This is preliminary work. In the future, we aim to evaluate its performance on real patient data. Furthermore, we hope that this method opens the door to a new, integrated acquisition-reconstruction pipeline, where the saved data is used to dynamically



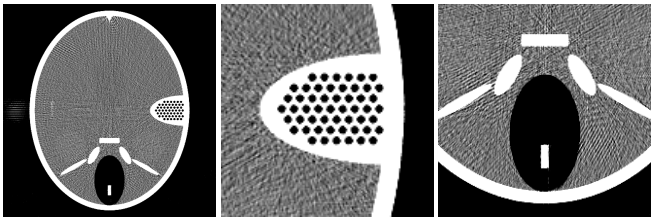
(a) uniform 1×1 binning, $r = 0\%$, SSIM(1)$= 1$



(b) uniform 2×2 binning, $r = 75\%$, SSIM(1)$= 0.773$



(c) dynamic 3×3 binning, $r = 75\%$, SSIM(1)$= 0.893$



(d) dynamic 5×5 binning, $r = 80\%$, SSIM(1)$= 0.934$

Fig. 3. Qualitative results on noisy phantom data. Simulation parameters and quantitative metrics are listed in Tab. 2. Window: C=60HU, W=500HU. See text for details.

vary the angular sampling to significantly improve the quality the reconstructed volumes.

Acknowledgements

This work was supported by the Research Training Group 1773 "Heterogeneous Image Systems", funded by the German Research Foundation (DFG).

6. REFERENCES

[1] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-Compensated and Gated Cone Beam Filtered Back-Projection for 3-D Rotational X-Ray Angiography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 898–906, 2006.

[2] A. de Sousa, "Smart Cameras as Embedded Systems," in *Proceedings First International Conference On Computer Applications*, 2003, pp. 105–112.

[3] A. N. Belbachir and P. M. Göbel, "Smart Cameras: A Historical Evolution," in *Smart Cameras*, pp. 3–17. Springer, 2010.

[4] M. Leeser, S. Miller, and H. Yu, "Smart Camera Based on Reconfigurable Hardware Enables Diverse Real-time Applications," in *12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*, 2004, pp. 17–155.

[5] A. Aichert, M. Berger, J. Wang, N. Maass, A. Doerfler, J. Hornegger, and A. Maier, "Epipolar Consistency in Transmission Imaging," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1–15, 2015.

[6] E.Y. Sidky, C.M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[7] X. Pan, E.Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?," *Inverse problems*, vol. 25, no. 12, pp. 1–50, 2009.

[8] A. Astrom, R. Forchheimer, and P. Ingelhag, "An Integrated Sensor/Processor Architecture Based on Near-Sensor Image Processing," in *Computer Architectures for Machine Perception*, 1993, pp. 17–154.

[9] L. Albani, P. Chiesa, D. Covi, G. Pedegani, A. Sartori, and M. Vatteroni, "VIsoc: a smart camera SoC," in *Solid-State Circuits Conference*, 2002, pp. 367–370.

[10] J. Dubois D., Ginhac, M. Paindavoine, and B. Heyrman, "A 10 000 fps CMOS sensor with massively parallel image processing," *Solid-State Circuits*, vol. 3, no. 3, pp. 706–717, 2008.

[11] A. Gabiger, R. Weigel, S. Oeckl, and P. Schmitt, "Enhancement of CT Image Quality via Bilateral Filtering of Projections," in *The first international conference on image formation in X-ray computed tomography*, 2010, pp. 140–143.

[12] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J.-H. Choi, C. Riess, A. Keil, and R. Fahrig, "CONRAD — A software framework for cone-beam imaging in radiology," *Medical Physics*, vol. 40, no. 11, pp. 111914, Nov. 2013.

[13] G. Lauritsch and H. Bruder, "FORBILD Head Phantom," `http://www.imp.uni-erlangen.de/forbild/english/results/head/head.html`, 1999.

[14] J. H. Hubbell and S. M. Seltzer, "Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients from 1 keV to 20 MeV for Elements Z = 1 to 92 and 48 Additional Substances of Dosimetric Interest," Tech. Rep., NIST, Physical Measurement Laboratory, July 2004, `http://www.nist.gov/pml/data/xraycoef/index.cfm`.

[15] B. Lorch, M. Berger, J. Hornegger, and A. Maier, "Projection and Reconstruction-Based Noise Filtering Methods in Cone Beam CT," in *Bildverarbeitung fr die Medizin*, 2015, pp. 59–64.

[16] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[17] Z Wang, A C Bovik, H R Sheikh, and E P Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.